DOCUMENT RESUME

ED 303 507 TM 012 754

AUTHOR Davidson, Betty M.
TITLE The Case against Using Stepwise Research Methods.
PUB DATE Nov 88
NOTE 11p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (17th, Louisville, KY, November 9-11, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Predictive Measurement; *Predictor Variables; Regression (Statistics); Researchers; *Research Methodology; Research Problems; *Statistical Analysis
IDENTIFIERS *Stepwise Regression

ABSTRACT

Researchers sometimes use stepwise methods to eliminate variables from analyses when the variables do not appreciably improve the ability to explain or predict inferences about the importance of various predictor variables. It is argued that stepwise methods are usually not appropriate for these purposes for three reasons. First, some researchers interpret stepwise results in which a given number of predictor variables have been selected as indicating that the predictor variables are the best variables to use if the predictor variable set is limited to the size of the given predictors ("g"). Stepwise methods add the next most helpful predictor at each forward step; however, this is not the same as finding the best predictors for a predictor variable of size "g". Second, if two predictor variables are closely correlated, there is a strong probability that one of the variables will absorb the majority of the other variable's predictor variable, and therefore be excluded. Third, most researchers use the wrong degrees of freedom to evaluate stepwise regression results. Small data sets are used to illustrate the fallacies that can be produced by incorrect use of stepwise methods. Two figures show the interrelationships of variables. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED303507

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Betty M. Davidson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

THE CASE AGAINST USING STEPWISE RESEARCH METHODS

Betty M. Davidson

University of New Orleans

Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY, November 9, 1988.

TM012754

ABSTRACT

Stepwise analytic methods are among the most commonly applied methods in educational and psychological research. These methods can be used in various analytic procedures, including multiple regression and discriminant function analysis. This paper argues that stepwise methods are usually not appropriate. Small data sets are presented to illustrate the fallacies that can be produced from the incorrect use of stepwise methods.

Researchers commonly use several predictor variables to estimate values on criterion variables. By using several predictor variables instead of one, researchers aim to reduce the errors of prediction to account for more of the criterion variable variance. Stepwise multiple regression analysis is a statistical procedure commonly used by researchers in this fashion (Kupper, Stewart, & Williams, 1976; Pope & Webster, 1972; Thompson, in press; Wilkinson, 1979). According to Marascuilo and Serlin (1988, p. 671), "The most popular method in use for selecting the fewest number of predictor variables necessary to guarantee adequate prediction is based on a model referred to as stepwise regression." Thompson (in press) concurs, noting that "Stepwise analytic methods may be among the most popular research practices employed in both substantive and validity research."

One of the problems in multiple regression that has received considerable attention is reducing the number of predictor variables in the prediction equation. In the construction of a regression equation it is not necessary or cost efficient to utilize each predictor variable in a given set. When the analysis begins with a large set of variables, the researcher may wish to eliminate those predictors which account for only an insignificant amount of the variation in the criterion variable (Kachigan, 1982).

The stepwise procedure uses the predictor variable that correlates most highly with the criterion variable as the

first entry. The second predictor variable entered is the one that accounts for most of the remaining or residual variance in the criterion variable. The forward stepwise procedure continues by adding the variable that accounts for the most variance in the criterion variable not already explained by the earlier variables, continuing until the inclusion of another variable would account for only an insignificant amount of variance in the criterion variable (Kupper et al., 1976; Pope et al., 1972; Thompson, in press). At each step each variable already entered is also subsequently considered for removal.

Much of the popularity of the stepwise procedure lies in the fact that this method provides a parsimonious account of the variance in the criterion variable. The procedure also provides a cost saving as it is impractical to collect data on a large set of variables when a smaller set would yield the same results.

The widespread use of the stepwise procedure has not been accompanied by a "widespread" understanding of the weakness in or misrepresentations arising from the procedure. As Wilkinson (1979, p. 168) notes, "Stepwise regression has a controversial role in statistical data analysis. Since the introduction of various automated techniques for selecting the 'best' subset of a set predictor variables in a multiple regression researchers have been warned about their indiscriminate use."

Researchers sometimes employ stepwise methods to eliminate variables from analyses when the variables do not appreciably

improve the ability to explain or predict inferences about the importance of various predictor variables. This paper argues that stepwise methods are usually not appropriate to use for these purposes, for three reasons.

First, as Thompson (in press) explains, some researchers interpret stepwise results in which g predictors variables have been selected as indicating that the predictor variables are the best variables to use if the predictor variables set is limited to size g. Stepwise methods add the next most helpful predictor at each forward step. Doing five steps identifies the five best predictors if the variables are added one at a time. This is not the same as finding the best predictors for a predictor variable set of size five!

Second, if two predictor variables are closely correlated, there is a strong probability that one of the variables will absorb the majority of the other variable's predictor variable, and therefore be excluded. Thus, according to Huberty (in press), "the order of variables entry in a stepwise analysis should not be used to assess relative variable contribution/importance." For example, predictor variables "X1" and "X2" might be closely correlated with each other but be independent of predictor variable "X3". The stepwise procedure might enter the predictor variables in the order "X1", then "X3", and then "X2". But, in fact, predictor variable "X2" could have a higher correlation with the criterion variable than "X3". The correlation of

predictor variable "X2" with the criterion variable was diminished due to the variable's close relationship with variable "X1". As Huberty (in press) notes, "This dependency or conditionality truly makes variable importance as determined by stepwise analysis very questionable."

Figures 1 and 2 illustrate this point. Given the relationships portrayed in Figure 1, the predictor variables would be entered in the order, "X1", "X2", and "X3". This order would, in fact, honor the predictive power of these three variables in this case. Given the Figure 2 situation, the variables would be entered in the order "X1", "X3", and "X2". The less informed researcher might conclude on this basis that "X2" is the worst of the three predictors.

Insert Figures 1 and 2 About Here

Third, most researchers employ the wrong degrees of freedom to evaluate stepwise regression results. Most researchers employ the F test to evaluate changes in the squared R when an additional variable is entered at a particular step. The common method for calculating degrees of freedom for such a test involves the use of the number of predictors in the numerator and (n - 1 minus the number of entered predictors) as the denominator of df error. For example, suppose for n = 100 people and q = 50 predictor variables that the researchers perform a two step

analysis. Most commonly the resulting degrees of freedom used to evaluate the squared R will be 2/ (100 - 1 - 2) = 2/97.

However, some researchers would argue that the degrees of freedom ought to be arrived at differently. They would maintain that the df error = (100 - 1 - 50) = 49, since the two predictor variables produced by the stepwise procedure were selected by consulting a much larger set of predictors. This more conservative approach would have the effect of making the df error smaller, the mean squared error larger, and the calculated F smaller. With the smaller calculated F the researcher might not be able to reject the null hypothesis.

Kerlinger (1986, p. 545) states that "the research problem and the theory behind the problem [and not stepwise methods] should determine the order of entry of variables in multiple regression analysis." Predictor variables should be selected based on the research problem. These variables should be entered all at once. Researchers who employ stepwise methods, particularly if they fail to use replication or cross-validation methods, should consider Cliff's (1987, pp. 120-121) argument that "a large proportion of the published results using this method probably present conclusions that are not supported by the data."

References

Cliff, N. (1987). Analyzing multivariate data. New York: Harcourt Brace Jovanovich.

Huberty, C. J. (in press). Problems with stepwise methods--better alternatives. In B. Thompson (Ed.), Advances in social sciences methodology (vol. 1). Greenwich, CT: JAI Press.

Kachigan, S. K. (1982). Multivariate statistical analysis. Radius Press: New York.

Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rinehart, and Winston.

Kupper, L. L., Stewart, J. R., & Williams, K. A. (1976). A note on controlling significance levels in stepwise regression. American Journal of Epidemiology, 103. 13-15.

Pope, P. T., & Webster, J. T. (1972). The use of an F-statistic in stepwise regression procedures. Technometrics, 14, 327-340.

Marascuilo, L. A., & Serlin, R. C. (1988). Statistical methods for the social and behavioral sciences. New York: W. H. Freeman.

Thompson, B. (in press). Why won't stepwise methods die? Measurement and Evaluation in Counseling and Development.

Wilkinson, L. (1979). Tests of significance in stepwise regression. Psychological Bulletin, 86, 168-174.

Figure 1

Predictor Variables with No Correlation with Each Other

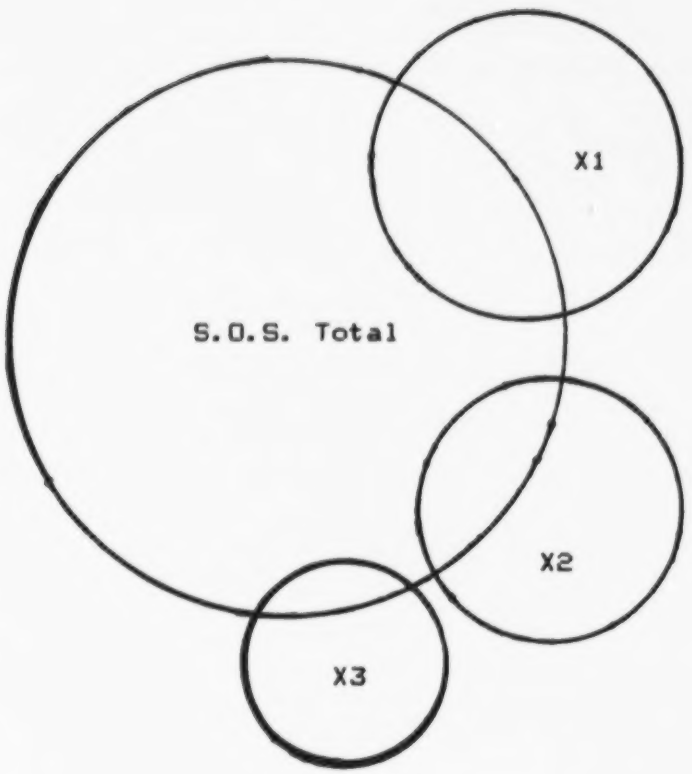

Figure 2

Correlating Predictor Variables